

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-161335

(43) 公開日 平成8年(1996)6月21日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27				
17/28		8420-5L	G 0 6 F 15/ 38	J
		8420-5L		T

審査請求 未請求 請求項の数 2 O L (全 13 頁)

(21) 出願番号 特願平6-298659

(22) 出願日 平成6年(1994)12月1日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 福持 陽士

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

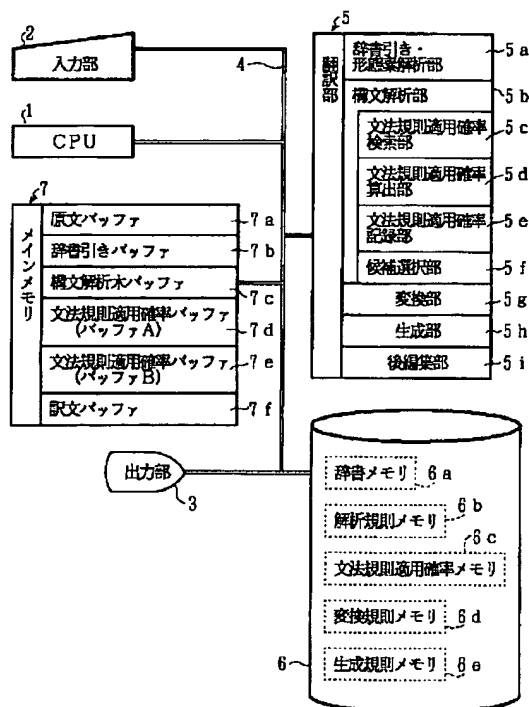
(74) 代理人 弁理士 原 謙三

(54) 【発明の名称】 機械翻訳装置

(57) 【要約】

【構成】 文法規則適用確率メモリ 6 c に格納された各文法規則の適用確率に基づいて構文構造の優先度を算出し、文法規則適用確率算出部 5 d が、ユーザが最適であると選択した翻訳文の基となった構文構造に適用された各文法規則の適用回数を、上記文法規則適用確率メモリ 6 c に格納された各文法規則の適用確率に反映させる。

【効果】 人間の翻訳者が構文構造の適切さに関して経験的に持っている知識を装置に学習させることができ、同一の原文から複数の構文構造が得られる場合に、最適な構文構造をより適切に選択することが可能となり、翻訳効率の向上を図ることができる。



1

【特許請求の範囲】

【請求項1】辞書および文法規則に基づいて原言語の文を解析して構文構造を求める構文解析手段と、上記構文構造から目標言語の翻訳文を生成する変換・生成手段と、生成された翻訳文を出力する出力手段とを備え、原言語の文書为目标言語に翻訳して出力する機械翻訳装置において、

各文法規則の適用確率を格納する記憶手段と、

上記構文解析手段が解析を行う際に適用した文法規則と、上記記憶手段に格納されている各文法規則の適用確率とに基づいて、各構文構造の優先度を算出する構文優先度算出手段と、

上記出力手段における翻訳文の出力順位を、各翻訳文の基となった構文構造の優先度に基づいて制御する制御手段と、

最も適切な翻訳文としてユーザが選択した翻訳文の基となった構文構造を上記構文解析手段が求める際に適用された各文法規則の適用回数を、上記記憶手段に格納されている各文法規則の適用確率に反映させる適用確率更新手段とを備えていることを特徴とする機械翻訳装置。

【請求項2】ユーザが文書の種類を入力する文種入力手段をさらに備えていると共に、

各文法規則の適用確率が文書の種類別に設けられて上記記憶手段に格納されており、上記文種入力手段から入力された文書の種類に基づいて、上記構文優先度算出手段および適用確率更新手段が、上記記憶手段に格納されている各文法規則の適用確率を選択することを特徴とする請求項1記載の機械翻訳装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、例えば、入力された英語の文章を日本語に翻訳して出力するような、自然言語を翻訳する機械翻訳装置等に関するものである。

【0002】

【従来の技術】従来、例えば、英語を日本語に翻訳する翻訳装置のように、ある言語（原言語）で書かれた文書を別の言語（目標言語）に翻訳するための機械翻訳装置が知られている。

【0003】上記の機械翻訳装置は、入力された原言語を、内部に記憶している辞書を用いて、単語等の形態素に分割すると共に、その形態素の品詞情報等を得て、原言語の文を構成する単語間の係り受けを解析して構文構造を決定し、上記構文構造を目標言語の構文構造に変換し、目標言語の構文構造から目標言語の表層文を生成することによって翻訳処理を行うものである。

【0004】しかし、人間の翻訳者であれば、単語の意味や前後関係等からほぼ経験的に判断することによって、原文に対する翻訳文を比較的容易に一意に決定することができるが、装置が翻訳を行う際には、ほとんどの場合において、原文に対する翻訳文は複数種類が作成さ

2

れ得る。これは、例えば、前記のように、品詞情報に基づいて単語間の係り受けを解析して原言語の構文構造を作成する際に、複数の品詞を有する単語が原言語の文中に存在する場合には、その単語と他の単語との係り受けを解析するために適用すべき文法規則が複数種類存在し、ひいては、作成される構文構造も複数種類になることによる。

【0005】このような場合に、原言語の文から作成され得るすべての構文構造に基づいて翻訳文を生成して出力することは、適切な翻訳からあまりにもかけ離れた翻訳文も出力されることとなり、ユーザがこれらの翻訳文から適切なものを選択する手間を増大させるので、翻訳効率の点からも好ましいことではない。

【0006】これに対して、例えば、特開平6-195377号公報には、同一の原文から複数種類の翻訳文が得られ、その内の1個の翻訳文を最適な翻訳文としてユーザが選択した場合に、上記の複数種類の翻訳文のそれぞれ基となっている複数種類の構文構造を互いに比較し、ユーザに選択された翻訳文の基となった構文構造の特徴点をルール化して記憶手段に記憶することによって、最適とされる構文構造を学習し、同一の原文から複数の構文構造が得られた場合には、過去に上記のように学習したルールに基づいて、複数種類の翻訳文の出力順位等を決定する構成が開示されている。さらに、複数の構文どうしが競合した場合に、他の構文構造に対してどの程度の強さで優先するかを示す優先度を各構文構造に対して定め、この優先度に従って適用すべき構文構造を決定する方法が提案されている。

【0007】

【発明が解決しようとする課題】しかしながら、上記公報に開示された構成では、ユーザが最適と考える構文構造が、ユーザの選択によって記憶手段に一旦記憶されると、当該構造の出現頻度に関係なく優先すべき構文構造として学習されてしまうため、学習の内容が一般的なものでなかった場合、かえって適切でない翻訳文を出力する結果となるという問題点を有している。また、上記公報に開示された構成では、上記のように優先度を用いた構文構造決定の方法を提案した段階にとどまり、他の構文構造に対する優先度を各構文構造に対してどのように設定するかについては技術的に未解決のままである。

【0008】本発明の目的は、原文から得られる複数の構文構造に基づいた翻訳文を出力する際に、優先して出力すべき翻訳文をより適切に決定できる機械翻訳装置を実現することである。

【0009】

【課題を解決するための手段】上記の目的を達成するために、本発明の請求項1記載の機械翻訳装置は、辞書および文法規則に基づいて原言語の文を解析して構文構造を求める構文解析手段と、上記構文構造から目標言語の翻訳文を生成する変換・生成手段と、生成された翻訳文

を出力する出力手段とを備え、原言語の文書为目标言語に翻訳して出力する機械翻訳装置において、各文法規則の適用確率を格納する記憶手段と、上記構文解析手段が解析を行う際に適用した文法規則と、上記記憶手段に格納されている各文法規則の適用確率とに基づいて、各構文構造の優先度を算出する構文優先度算出手段と、上記出力手段における翻訳文の出力順位を、各翻訳文の基となった構文構造の優先度に基づいて制御する制御手段と、最も適切な翻訳文としてユーザが選択した翻訳文の基となった構文構造を上記構文解析手段が求める際に適用された各文法規則の適用回数を、上記記憶手段に格納されている各文法規則の適用確率に反映させる適用確率更新手段とを備えていることを特徴としている。

【0010】請求項2記載の機械翻訳装置は、請求項1記載の機械翻訳装置において、ユーザが文書の種類を入力する文種入力手段をさらに備えていると共に、各文法規則の適用確率が文書の種類別に設けられて上記記憶手段に格納されており、上記文種入力手段から入力された文書の種類に基づいて、上記構文優先度算出手段および適用確率更新手段が、上記記憶手段に格納されている各文法規則の適用確率を選択することを特徴としている。

【0011】

【作用】請求項1記載の構成によれば、入力された原言語の文を構文解析手段が解析する際に適用した文法規則と、記憶手段に格納されている各文法規則の適用確率とに基づいて、構文優先度算出手段が各構文構造の優先度を算出し、例えば、複数の構文構造の優先度を互いに比較して一番優先度が高い構文構造から先に出力する等のように、出力手段における翻訳文の出力順位を制御手段が制御する。さらに、適用確率更新手段が、最も適切な翻訳文としてユーザが選択した翻訳文の基となった構文構造を上記構文解析手段が求める際に適用した各文法規則の適用回数を、記憶手段に格納されている各文法規則の適用確率に反映させる。

【0012】これにより、1つの原文から得られる複数種類の翻訳文から最も適切な翻訳文を選択する際のユーザの判断が、文法規則の適用確率に反映されて記憶手段に格納される。このため、翻訳処理を行う度に、構文構造の適切さに関して人間の翻訳者が経験的に持っている知識が蓄積的に装置に学習されることとなる。この結果、複数種類の翻訳文から、最も適切な翻訳文が優先的に選択されて出力される可能性が高くなり、翻訳処理の効率の向上を図ることができる。

【0013】請求項2記載の構成によれば、構文優先度算出手段および適用確率更新手段が、文書の種類別に記憶手段に格納されている各文法規則の適用確率から、文種入力手段からユーザが入力した文書の種類に対応する適用確率を選択する。

【0014】これにより、例えば、技術マニュアルや新聞記事等のように、頻繁に使用される構文構造が異な

ていると思われる複数種類の文書を同一の翻訳装置で翻訳する場合においても、各文書の種類に応じて適切な翻訳文が優先的に出力される可能性が高くなり、翻訳処理の効率の向上を図ることができる。

【0015】

【実施例】

【実施例1】本発明の一実施例について図1ないし図9、および図12に基づいて説明すれば、以下の通りである。なお、本実施例では、原言語が英語であり、目的言語が日本語であるような機械翻訳装置を例にあげて説明する。

【0016】本実施例における機械翻訳装置は、図1に示すように、以下の各部の動作を制御する制御手段としてのCPU（中央処理装置）1と、RAM等からなるメインメモリ7と、原言語の単語列を入力するためのアルファベットキーや各種命令を入力するためのコマンドキー等を備えたキーボード等の入力部2（文種入力手段）と、入力された単語列を原言語から目的言語へ翻訳する処理を行う翻訳部5と、辞書および文法規則等を記憶しているハードディスク6と、目的言語に翻訳された文等を出力するための、例えばCRT（陰極線管）やLCD（液晶表示装置）等で構成される出力部3とを備えている。上記の入力部2、翻訳部5、ハードディスク6、メインメモリ7および出力部3は、データや制御信号等を伝達するバスライン4によって互いに接続されており、上記CPU1の制御に基づいて動作する。

【0017】上記ハードディスク6は、辞書を格納している辞書メモリ6aと、原文を解析して構文構造を求める際に適用すると、上記文法規則のそれぞれの適用確率を格納している文法規則適用確率メモリ6c（記憶手段）と、原文から得られた構文構造を目的言語の構文構造に変換する際に適用する変換規則を格納している変換規則メモリ6dと、目的言語の構文構造から目的言語の文を生成する際に適用する生成規則を格納している生成規則メモリ6eとを含んでいる。

【0018】上記翻訳部5は、同図に示すように、原言語の文に含まれる各単語の訳語および品詞情報等を得るための辞書引き・形態素解析部5aと、上記品詞情報および文法規則に基づいて、構文構造、すなわち各単語間の係り受け関係、を示す構文解析木を得るための構文解析部5bと、入力された原文に対する構文解析木の構造を目的語に対する構文解析木の構造に変換する変換部5gと、最終的に目標言語の文を生成する生成部5hと、上記入力部2に出力された翻訳結果をユーザの指示にしたがって修正する後編集部5iとを含んでいる。

【0019】上記構文解析部5bは、ハードディスク6の文法規則適用確率メモリ6cを検索する文法規則適用確率検索部5cと、文法規則の適用確率を計算する文法規則適用確率算出部5d（適用確率更新手段）と、算出された適用確率をハードディスク6の文法規則適用確率

メモリ6cへ格納する文法規則適用確率記録部5eと、文法規則適用確率に基づいて各構文解析木の優先度、すなわち各構文解析木の適切さを算出する候補選択部5f（構文優先度算出手段）とによって構成されている。

【0020】ここで、「I have both pens and pencils.」という英文を日本語文に翻訳する際の動作について、図1ないし図3を参照しながら説明する。

【0021】まず、入力部2から上記の英文が入力され、CPU1の制御の下でメインメモリ7の原文バッファ7aへ一旦格納される。翻訳部5は、CPU1の制御の下で、原文バッファ7aから上記英文を取り出し、辞書引き・形態素解析部5aによってハードディスク6の辞書メモリ6aを参照し、各単語の訳語と品詞情報とを得てメインメモリ7の辞書引きバッファ7bに格納する。

【0022】この時、辞書メモリ6aから得られる各単語に対する品詞情報は、以下の通りとなる。すなわち、「I」は名詞（pron）、「have」は動詞（verb）、「both」は限定詞（det）と接続詞（conj）と名詞（pron）との内のいずれかであり、「pens」は名詞（pron）、「and」は接続詞（conj）、「pencils」は名詞（pron）、「.」は終止符（end）である。

【0023】次に、翻訳部5の構文解析部5bが、上記のように得られた品詞情報と、ハードディスク6の解析規則メモリ6bに格納されている文法規則とに基づいて、各単語間の係り受け関係を示す構文解析木を求めてメインメモリ7の構文解析木バッファ7cに格納する。

【0024】この構文解析木は、次のようにして求められる。すなわち、解析規則メモリ6bには、図2に示す文法規則が格納されている。同図に示すように、各文法規則には、1から19までの文法規則番号が付けられており、例えば、一番上の行に示されている文法規則番号が1の文法規則（以下、規則R1のように表記する）は、『文章（SS）は、文（S）と終止符（end）とから成っている。』ということを表わす規則である。

【0025】なお、同図において、規則R19の下部に示すように、これらの文法規則中に例えば「NP」、「VP」等のように大文字で示されている記号は、非終端記号と呼び、同図中に示す文法規則に従ってさらに解析することが可能な構造であることを示している。一方、例えば「noun」、「conj」等のように小文字で示されている記号は、終端記号と呼び、これ以上は解析できない品詞等に対応している。つまり、非終端記号を、同図に示す文法規則に従って終端記号へと分解していくことにより、構文解析木が求められるわけである。

【0026】上記の英文を図2に示す文法規則に従って解析した結果の構文解析木の一例は、図3に示すようになる。同図中で大文字で示す非終端記号の横のかっこ内には、その非終端記号を解析するのに用いた文法規則を示している。例えば、同図の一番上に示した文章（SS）

は、図2に示す規則R1に従って、文（S）と終止符（end）とに分解され、上記文章（SS）の左下に示した文（S）は、規則R2に従って、名詞句（NP）と動詞句（VP）とにさらに分解されたことを示す。

【0027】このようにして、すべての非終端記号が、図3に略楕円形で囲んで示す終端記号に分解された様子を示す構文解析木は、構造解析木バッファ7cに一旦格納される。その後、変換部5gは、構文解析木バッファ7cから構文解析木を取り出して、取り出した構文解析木を、日本語の文法規則に従って日本文に対する構文解析木に変換する。続いて、生成部5hが、適切な助詞「は」等を補うことにより、図3の下部に示すような、「私は、両方のペンと鉛筆を持っている。」という翻訳文を求めて、メインメモリ7の訳文バッファ7fに格納する。

【0028】以上のような動作によって、上記の英文の翻訳文が求められたわけであるが、「both」という単語が複数の品詞を持つことから、上記の英文に対する構文解析木として、前記の図3に示した構文解析木の他に、図4に示すような構文解析木を得ることができる。このように、原文に対する翻訳文は必ずしも1種類であるとは限らず、複数の候補から適切な翻訳文を選択することが必要となる。

【0029】以下に、上記のように複数種類の翻訳文の候補が存在する場合の、最適な翻訳文を選択するための動作について、図5のフローチャートを参照しながら説明する。

【0030】まず、翻訳対象となる原文が入力部2から入力される（ステップ1、以下、S1と表記する）。あるいは、入力されて一旦ハードディスク6へ格納されていた原文を読み出してもよい。

【0031】続いて、ハードディスク6の文法規則適用確率メモリ6cに格納された文法規則適用確率を、メインメモリ7の文法規則適用確率バッファ7dのバッファAへ読み込む（S2）。

【0032】上記の文法規則適用確率とは、非終端記号を解析する際に、その非終端記号に適用可能な文法規則すべてに対して、ある文法規則が適用される確率を示すものである。すなわち、図2に示すように、例えば、非終端記号NPに適用される文法規則は、規則R3ないし規則R8の6種類であるので、図6に示すように、規則R3ないし規則R8の文法規則適用確率は $1 \div 6 \approx 0.16$ となる。また、非終端記号NP1に適用される文法規則である規則R9および規則R10の文法規則適用確率はそれぞれ $1 \div 2 = 0.50$ であり、その他の文法規則に対する確率も同様に図6に示すように求めることができる。

【0033】ただし、上記の図6に示された文法規則適用確率は、各文法規則に優先度が付けられていない初期状態における確率であり、本実施例の文法規則適用確率

は、以降に説明するように、翻訳結果をユーザが選択する度にその選択を学習して優先度が付けられるように更新されてゆく。

【0034】ユーザから翻訳の実行指示が入力されると(S3)、原文の1文がメインメモリ7の原文バッファ7aにセットされる(S4)。なお、ここで、上記の原文の1文として、前記の英文「I have both pens and pencils.」がセットされたものとする。

【0035】次に、原文バッファ7aに読み込まれた原文の各単語に対して、辞書引き・形態素解析部5aがハードディスク6の辞書メモリ6aを参照することにより、訳語や品詞情報を取得して辞書引きバッファ7bへ格納する(S5)。

【0036】上記S5に続く、図5で点線で囲んで示すS6ないしS10は、図1に示す構文解析部5bにおいて行われる処理である。まず、前述したように、上記英文に対する構文解析木を求める処理が行われ、求められた図3および図4に示す2種類の構文解析木は、構文解析木バッファ7cへ格納される(S6)。

【0037】ここで、上記の2種類の構文解析木のどちら

$$\begin{aligned} P_a &= P(R1) \times P(R2) \times P(R7) \times P(R12) \times P(R17) \times \\ &\quad P(R6) \times P(R3) \times P(R9) \times P(R4) \times P(R9) \\ &= 1 \times 1 \times 0.16 \times 0.25 \times 0.33 \times \\ &\quad 0.16 \times 0.16 \times 0.50 \times 0.16 \times 0.50 \\ &= 0.000013516 \end{aligned}$$

となる。

※度を P_a とすると、

【0039】同様にして、図4に示す構文解析木の優先※

$$\begin{aligned} P_b &= P(R1) \times P(R2) \times P(R7) \times P(R12) \times P(R17) \times \\ &\quad P(R5) \times P(R4) \times P(R9) \times P(R4) \times P(R9) \\ &= 1 \times 1 \times 0.16 \times 0.25 \times 0.33 \times \\ &\quad 0.16 \times 0.16 \times 0.50 \times 0.16 \times 0.50 \\ &= 0.000013516 \end{aligned}$$

となる。

【0040】続くS8において、上記のように求められた複数の構文解析木の優先度を比較して、最も優先度が高い構文解析木を最良候補として選択するが、上記の場合のように、複数の構文解析木の優先度が同じ値になった場合には、例えば、次のように、品詞の優先度に基づいて構文解析木の選択を行ってもよい。つまり、複数の品詞を有する単語に対して、各品詞間の優先度を数値で示す品詞優先度を規定して辞書メモリ6aに格納しておく、文中のすべての単語の品詞優先度の加算合計値が小さい方の構文解析木を選択する方法である。なお、優先度が最も高い品詞に対する品詞優先度を1として、以下優先度が低くなる順に2・3…と規定されるものとする。また品詞が1種類しかない単語に対しては、品詞優先度は1と規定されるものとする。

【0041】上記英文の場合には、「both」が3種類の品詞を有しており、限定詞(det)、接続詞(conj)、名詞(pron)の順に優先度が付けられて、図12に示す

*らを選択するかを決定するために、文法規則適用確率検索部5cが、上記S2でバッファAに読み込まれた文法規則適用確率を検索し、さらに候補選択部5fが、各構文解析木の優先度を、各構文解析木に適用されている文法規則の文法規則適用確率に基づいて算出する(S7)。すなわち、各構文解析木で用いられた文法規則の文法規則適用確率すべての積を、その構文解析木の優先度とする。

【0038】例えば、図3に示す構文解析木の優先度は、以下のようにして求めることができる。図3に示す構文解析木で用いられている文法規則は、同図の非終端記号の横の括弧内内で示されているように、規則R1、規則R2、規則R7、規則R12、規則R17、規則R6、規則R3、規則R9、規則R4および規則R9である。ここで、例えば規則R1の文法規則適用確率を $P(R1)$ のように表すとすると、上記構文解析木の優先度 P_a は、まだユーザによって翻訳結果が選択される前の初期状態であるので、図6に示す初期状態の文法規則適用確率から、以下のように算出される。すなわち、

ように、それぞれの品詞優先度が1・2・3と規定されているとすると、上記図3の構文解析木においては、「both」は限定詞(det)であり、この品詞優先度は1であるため、全単語の品詞優先度の加算合計値は、7となる。一方、図4に示す構文解析木においては、「both」は接続詞(conj)であり、品詞優先度は2と規定されているために、全単語の品詞番号の加算合計値は、8となる。この結果、図3に示す構文解析木の方が優先度が高いとして、翻訳を行うための最良候補として選択される。

【0042】次のS9では、文法規則適用確率算出部5dにおいて、上記S8で選択された最良候補の構文解析木における文法規則適用確率が算出される。この場合に、最良候補として選択された、図3で示す構文解析木で用いられた文法規則とその回数は、非終端記号SSに対しては規則R1が1回、非終端記号Sに対しては規則R2が1回、非終端記号NPに対しては、規則R3、規則R4、規則R6および規則R7が1回ずつで合計4回、非

終端記号NP1 に対しては規則R 9 が2回、非終端記号VP に対して規則R 1 2 が1回、非終端記号V に対して規則R 1 7 が1回である。

【0043】これらから、図7に示すように、各非終端記号に対して適用された文法規則のそれぞれの文法規則適用確率が算出される。例えば、非終端記号NPに対して適用される規則R 3 ないしR 8 については、上記で用いられた規則R 3、規則R 4、規則R 6 および規則R 7 に対して $1 \div 4 = 0.25$ がそれぞれ算出されて設定される。また、非終端記号NP1 に対して適用される規則R 9 およびR 1 0 については、規則R 9 のみに対して $2 \div 2 = 1$ が算出されて設定される。

【0044】すなわち、各規則の文法規則適用確率は、上記S 8 で選択された構文解析木における各規則の使用回数を、その規則が適用される非終端記号の出現回数の総数で割ることによって算出される。

【0045】続くS 1 0 では、文法規則適用確率記録部5 e の制御により、上記S 9 で算出された文法規則適用確率が、文法規則適用確率バッファ7 d におけるバッファBに格納される。

【0046】さらに、次のS 1 1 では、上記S 8 で最良候補として選択された図3の構文解析木が、日本語の文法に基づいて、日本語の構文解析木に変換され、さらに、続くS 1 2 で、助詞等が補われて翻訳文が生成される。そして、S 1 3 で、生成された翻訳文「私は、両方のペンと鉛筆を持っている。」が出力部3に表示されることとなる。

【0047】ユーザは、表示されたこの翻訳文を見て、正しい翻訳がなされたかどうかを判断し(S 1 4)、この場合は、正しい翻訳ではないので、次候補選択の指示を入力する(S 1 5)。上記指示が入力されると前記S 8 へ再び戻って、さきほど選択した構文解析木の次に優先度が高い構文解析木を、新たな最良候補として選択する。この場合、図4に示す構文解析木が新たな最良候補として選択されることとなる。

【0048】再び、S 9 において、上記の図4に示す構文解析木に基づいて、文法規則適用確率の算出が行われ、S 1 0 でバッファBへの上記文法規則適用確率の格納が行われる。この時、前回のS 9 の処理で図3の構文解析木から求められてバッファBに格納されていた文法規則適用確率はすべて消去されるか上書きされて、バッファBの内容は、図4の構文解析木から求められた確率等が、図8に示すとおりに格納されている状態となる。

【0049】さらに、S 1 1 ないしS 1 2 の処理が行われ、S 1 3 において、「私は、ペンと鉛筆との両方を持っている。」という翻訳文が表示される。この翻訳文はユーザが所望の正しい翻訳文であるため、S 1 4 からS 1 6 へ移行する。

【0050】S 1 6 では、文法規則適用確率が修正される。つまり、最終的にユーザに選択された翻訳文の基と

なった構文解析木を求める際に適用された各文法規則の使用頻度を、前記S 2 でバッファAに読み込まれた初期状態の文法規則適用確率に反映させるために、次のような計算を行う。

【0051】まず、図6に示すようにバッファAに格納されている初期状態の各文法規則の適用回数(この場合は、すべて1)に、図8に示すようにバッファBに格納されている各文法規則の適用回数をそれぞれ加算する。同様に、図6に示す初期状態における該当非終端記号の出現回数に、図8に示すバッファBに格納されている該当非終端記号の出現回数をそれぞれ加算する。

【0052】例えば、規則R 4 について見ると、図6に示す初期状態での適用回数は1回であり、図8に示すバッファBに格納されている適用回数は2回である。これを加算すると、図9に示すように、適用回数の合計は、3回になる。また同様に、該当非終端記号の出現回数の合計は、 $6 + 4 = 10$ 回となる。

【0053】このようにして、すべての文法規則について、初期状態の適用回数および該当非終端記号の出現回数と、バッファBに格納されている適用回数および該当非終端記号の出現回数とをそれぞれ加算した結果は、図9に示すようになり、さらに、各文法規則の適用回数を、該当非終端記号の出現回数で割ることによって、各文法規則に対する文法規則適用確率が同図に示すとおりそれぞれ算出される。例えば、上記の規則R 4 については、 $3 \div 10 = 0.30$ となる。

【0054】この結果、図6に示した初期状態では、同一の非終端記号に適用させる文法規則の適用確率は、同図から明らかなようにすべて等しくなっていたが、上記のように、文法規則適用確率を算出した結果は、図9に示すように、ユーザが選択した翻訳結果に適用された文法規則の適用確率が、他の文法規則の適用確率に比較して高くなっている。つまり、規則R 3 ないしR 8 を見てみると、ユーザが選択した翻訳結果の基となった、図4に示す構文解析木で適用された文法規則であるR 4、R 5 およびR 7 の文法規則適用確率が高いことは明らかである。さらに、R 4、R 5 およびR 7 の間でも、適用された回数が多かったR 4 が、他の2つの文法規則に比較して高い適用確率となっていることがわかる。

【0055】上記のように算出された文法規則適用確率は、バッファAに格納される。なお、この格納の際に、バッファAに予め格納されていた初期状態の文法規則適用確率は、格納に先立って消去されるか、あるいは、上書きされるので、バッファAには更新された文法規則適用確率のみが格納された状態となる。

【0056】続くS 1 7 では、入力された原文のすべてが翻訳されたか否かが判定され、入力された原文のすべての翻訳が終了したと判定された場合には、次のS 1 8 へ移行し、S 1 8 において、バッファAに格納されている各文法規則の適用回数、該当非終端記号の出現回数お

よび文法規則適用確率を、ハードディスク6へ格納し、翻訳処理を終了する。

【0057】一方、上記S17で、翻訳すべき原文がまだ残っていると判定された場合は、S4へ戻り、次の文が原文バッファ7aにセットされて、S5およびS6の処理が上記と同様に行われる。ここで、上記の英文「I have both pens and pencils.」が再び原文バッファ7aにセットされ、S5およびS6の処理によって、前記*

$$\begin{aligned} P_a &= P(R1) \times P(R2) \times P(R7) \times P(R12) \times P(R17) \times \\ &\quad P(R6) \times P(R3) \times P(R9) \times P(R4) \times P(R9) \\ &= 1 \times 1 \times 0.20 \times 0.40 \times 0.50 \times \\ &\quad 0.10 \times 0.10 \times 0.75 \times 0.30 \times 0.75 \\ &= 0.0000675 \end{aligned}$$

となり、同様にして、図4に示す構文解析木の優先度を※ ※P_bとすると、

$$\begin{aligned} P_b &= P(R1) \times P(R2) \times P(R7) \times P(R12) \times P(R17) \times \\ &\quad P(R5) \times P(R4) \times P(R9) \times P(R4) \times P(R9) \\ &= 1 \times 1 \times 0.20 \times 0.40 \times 0.50 \times \\ &\quad 0.20 \times 0.30 \times 0.75 \times 0.30 \times 0.75 \\ &= 0.000405 \end{aligned}$$

となる。

【0058】次のS8でこの結果を比較すると、P_aよりもP_bの方が大きいため、構文解析木の最良候補として図4に示す構文解析木が選択され、さらにS9ないしS12の処理を経て、S13で翻訳文「私は、ペンと鉛筆の両方を持っている。」が出力される。

【0059】上記の説明から明かなように、本実施例の翻訳装置は、原文から翻訳文を作成して出力し、最適な翻訳文をユーザに選択させ、ユーザが選択した翻訳文の基となった構文解析木に基づいて、各文法規則の適用確率を算出し、1文の翻訳処理を終了する度に文法規則適用確率を更新する構成である。これにより、ユーザが構文構造の適切さに関して経験的に持っている知識が文法規則適用確率として装置に学習されることとなる。

【0060】この結果、同一の原文から複数の構文構造が得られた場合に、ユーザの使用状況から学習した優先度に従って選択された翻訳文が、最良候補の翻訳文として逐次出力されるので、ユーザが所望する翻訳文が迅速に得られる確率を高めることができる。すなわち、翻訳効率を向上させることが可能となる。

【0061】なお、本実施例では、英語の文書を日本語に翻訳する装置を例に挙げて説明したが、本発明はこれに限定されるものではないことは言うまでもなく、また、文法規則についても本実施例に記載したものに限られるものではない。

【0062】〔実施例2〕本発明の他の実施例について図1、図5、図10および図11に基づいて説明すれば、以下の通りである。尚、説明の便宜上、前記の実施例1で説明した構成と同一の機能を有する構成には、同一の符号を付記し、その説明を省略する。

【0063】例えば、技術マニュアル、新聞記事あるい

*と同様に、図3および図4に示す構文解析木が求められたとする。この場合、続くS7で参照されるバッファAには、前述のように、図9に示すとおりに更新された文法規則適用確率が格納されており、S7で、この文法規則適用確率に基づいて各構文解析木の優先度を計算すると、次のようになる。すなわち、まず、図3に示す構文解析木の優先度をP_aとすると、

20 は小説等のように、異なる種類の文書においては比較的良好に使用される文体が異なるために、頻繁に使用される構文構造がそれぞれ異なるということが考えられる。このため、本実施例の翻訳装置は、翻訳処理の実行指示の入力に先立って、ユーザに対して、図1に示す入力部2から文書の種類（以下、文種と表記する）を入力するように指示を行う。また、同図に示すハードディスク6内には、文法規則適用確率メモリ6cが、複数種類の文種に対応するように設けられており、図11に示すように、各文種それぞれに対して、各文法規則の適用回数、各非終端記号の出現回数および文法規則適用確率が格納されている。

30 【0064】以下、図10に示すフローチャートを参照しながら、本翻訳装置の動作について説明する。なお、前記実施例1で説明した図5に示すフローチャートの各ステップと同じ処理を行うステップについては、同じステップ番号を付けて、その詳細な説明を省略する。なお、図5および図10を比較することで明かなように、図10に示すフローチャートは、図5におけるS1の後にS1'を行い、S2の代わりにS2'、S18の代わりにS18'を行うようになっている。

40 【0065】まず、翻訳すべき原文を入力する（S1）。次に、文種の指定を行うようにユーザに指示するメッセージを出力部3へ出力し、入力部2から入力された文種を読み込む（S1'）。ここで、例えば、「1. 文種指定なし」、「2. マニュアル」および「3. 新聞」の3種類の文種から選択を行うようにユーザに指示が出されたとする。なお、「1. 文種指定なし」は、特に文種を指定しない文章を翻訳する場合、「2. マニュアル」は、技術マニュアル等を翻訳する場合、「3. 新聞」は、新聞記事を翻訳する場合にそれぞれ選択される

ものとする。

【0066】上記S1'で文種が指定されると、続くS2'では、指定された文種に該当する文法規則適用確率メモリ6cから、各文法規則の適用回数、各非終端記号の出現回数および文法規則適用確率が読み出されて文法規則適用確率バッファ7dに格納される。

【0067】その後、S3で翻訳処理の実行指示がユーザによって入力されると、S4ないしS16の処理が前記実施例1と同様に行われる。さらに、S17で、翻訳すべき原文すべての処理が終了したと判定された場合は、上記S4ないしS16の処理の間に更新されてバッファAに格納されている、各文法規則の適用回数、各非終端記号の出現回数および文法規則適用確率を、該当する文種に対応する文法規則適用確率メモリ6cへ格納する。

【0068】上記のように、本実施例の翻訳処理装置は、複数種類の文種にそれぞれ対応した文法規則適用確率メモリ6cを備えており、翻訳処理の際に、文種を指定すると、指定された文種に対応した上記文法規則適用確率メモリ6cから、文法規則適用確率等を読み出し、これに基づいて翻訳処理を行い、さらに、翻訳処理を行う間に、前記実施例1と同様に各文法規則をユーザが適用する頻度を学習する構成となっている。

【0069】これにより、例えば、技術マニュアルや新聞記事等のように、頻繁に使用される構文構造が異なっていると思われる複数種類の文書を同一の翻訳装置で翻訳する場合においても、あらかじめ文種を指定することによって、指定された文種に対応する文法規則の適用確率に基づいてより適切な構文構造が決定されて優先的に出力されるため、翻訳処理の効率の向上を図ることができると共に、文書の種類に応じた適切な翻訳処理を行うことができる。

【0070】なお、本実施例では文章の種類を上記の3種類としたが、文章の種類はこれに限定されるものではなく、例えばユーザの希望に合わせて任意の種類に設定することが可能である。また、本装置の製作時にあらかじめ所定の複数種類の文種を設定しておくこともできるし、装置を購入した後に、ユーザが自分で設定を行う構成としてもよい。このような構成によれば、ユーザが翻訳する頻度が高い文種を自分で設定することができるため、無駄な設定を省き、装置の操作性を向上させることができる。

【0071】また、上記した各実施例においては、ユーザが翻訳文の選択を行って1つの文の翻訳処理が終了する度に、文法規則の適用確率が更新される例について説明したが、本発明はこれに限定されるものではなく、例えば、上記適用確率を更新するか否かをユーザが設定できる構成としてもよい。

【0072】

【発明の効果】以上のように、本発明の請求項1記載の

機械翻訳装置は、各文法規則の適用確率を格納する記憶手段と、上記構文解析手段が解析を行う際に適用した文法規則と、上記記憶手段に格納されている各文法規則の適用確率とに基づいて、各構文構造の優先度を算出する構文優先度算出手段と、上記出力手段における翻訳文の出力順位を、各翻訳文の基となった構文構造の優先度に基づいて制御する制御手段と、最も適切な翻訳文としてユーザが選択した翻訳文の基となった構文構造を上記構文解析手段が求める際に適用された各文法規則の適用回数を、上記記憶手段に格納されている各文法規則の適用確率に反映させる適用確率更新手段とを備えている構成である。

【0073】これにより、翻訳処理を行う度に、構文構造の適切さに関してユーザが経験的に持っている知識が蓄積的に装置に学習されることとなるため、複数種類の翻訳文から、最も適切な翻訳文が優先的に選択されて出力される可能性が高くなり、翻訳処理の効率の向上を図ることができるという効果を奏する。

【0074】請求項2記載の機械翻訳装置は、ユーザが文書の種類を入力する文種入力手段をさらに備えていると共に、各文法規則の適用確率が文書の種類別に設けられて上記記憶手段に格納されており、上記文種入力手段から入力された文書の種類に基づいて、上記構文優先度算出手段および適用確率更新手段が、上記記憶手段に格納されている各文法規則の適用確率を選択する構成である。

【0075】これにより、頻繁に使用される構文構造が異なっていると思われる複数種類の文書を同一の翻訳装置で翻訳する場合においても、各文書の種類に応じて適切な翻訳文が優先的に出力される可能性が高くなり、翻訳処理の効率の向上を図ることができるという効果を奏する。

【図面の簡単な説明】

【図1】本発明の一実施例における機械翻訳装置の概略構成を示すブロック図である。

【図2】上記機械翻訳装置の構文解析部において、構文解析を行う際に適用される文法規則の組合せの一例を示す説明図である。

【図3】原文と、上記原文から求められる構文解析木および翻訳文の対応関係の一例を示す説明図である。

【図4】原文と、上記原文から求められる構文解析木および翻訳文の対応関係の他の例を示す説明図である。

【図5】上記機械翻訳装置における翻訳処理の流れを示すフローチャートである。

【図6】初期状態における、上記の文法規則それぞれの適用回数と、上記文法規則が適用される非終端記号の出現回数と、文法規則適用確率との対応関係を示す説明図である。

【図7】上記図3に示す構文解析木における、上記の文法規則それぞれの適用回数と、上記文法規則が適用され

る非終端記号の出現回数と、文法規則適用確率との対応関係を示す説明図である。

【図8】上記図4に示す構文解析木における、上記の文法規則それぞれの適用回数と、上記文法規則が適用される非終端記号の出現回数と、文法規則適用確率との対応関係を示す説明図である。

【図9】学習を行った構文解析木における、上記の文法規則それぞれの適用回数と、上記文法規則が適用される非終端記号の出現回数と、文法規則適用確率との対応関係を示す説明図である。

【図10】本発明の他の実施例における機械翻訳装置が行う翻訳処理の流れを示すフローチャートである。

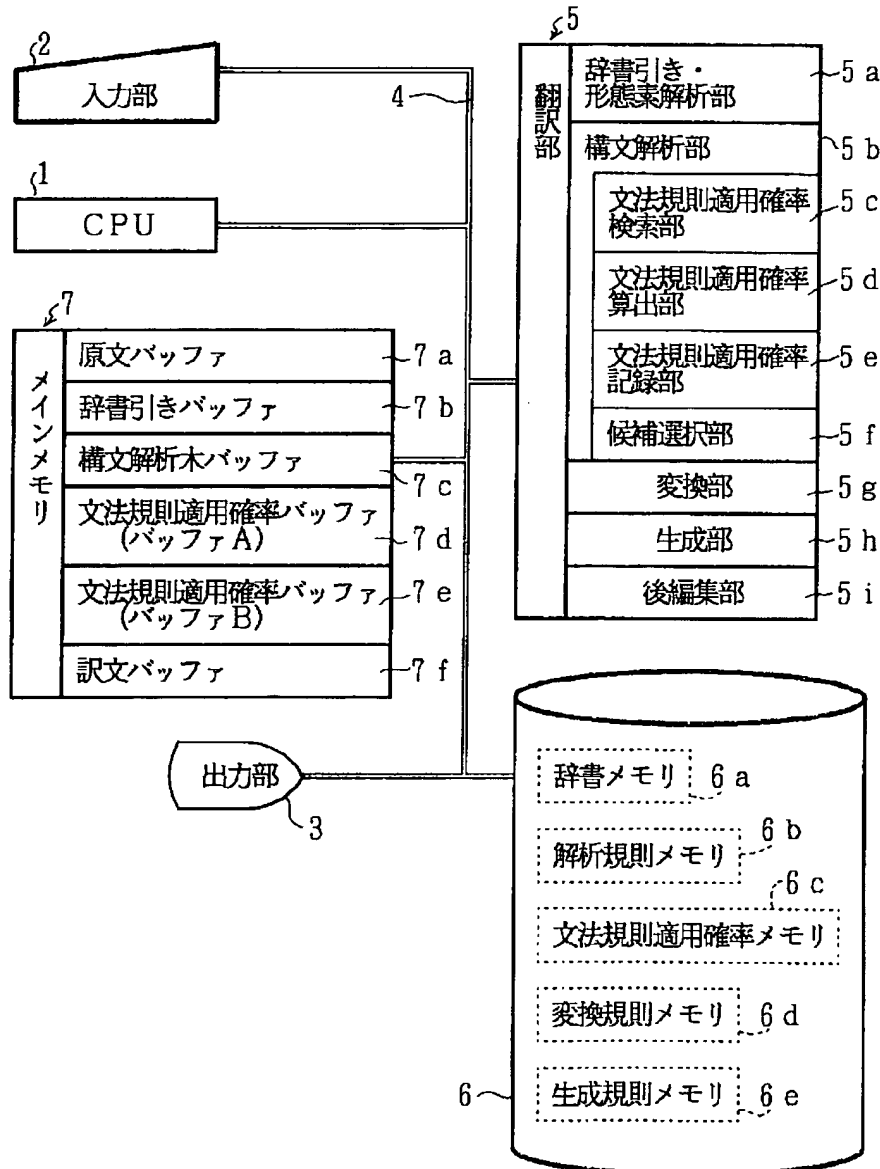
【図11】上記機械翻訳装置の構文解析部において適用される、文種別の文法規則適用確率等を示す説明図である。

【図12】各単語の品詞に付けられる優先度の一例を示す説明図である。

【符号の説明】

- 1 CPU (制御手段)
- 2 入力部 (文種入力手段)
- 5 f 候補選択部 (構文優先度算出部)
- 5 d 文法規則適用確率算出部 (適用確率更新手段)
- 6 c 文法規則適用確率メモリ (記憶手段)

【図1】



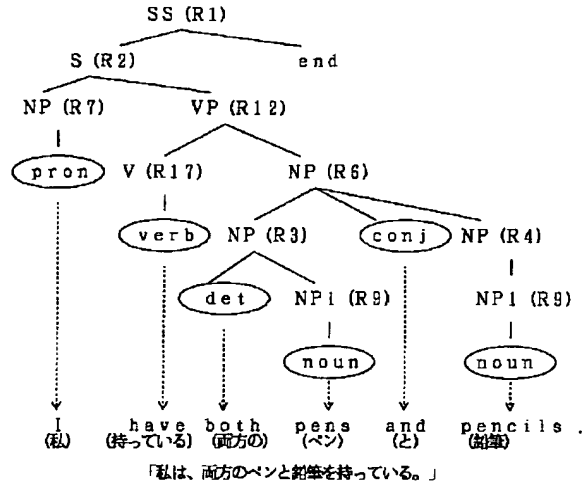
【図2】

- 1) SS → S end
- 2) S → NP VP
- 3) NP → det NP1
- 4) NP → NP1
- 5) NP → conj NP conj NP
- 6) NP → NP conj NP
- 7) NP → pron
- 8) NP → NP PP
- 9) NP1 → noun
- 10) NP1 → adj NP1
- 11) VP → V NP ADJP
- 12) VP → V NP
- 13) VP → V
- 14) VP → VP PP
- 15) PP → prep NP
- 16) ADJP → adj
- 17) V → verb
- 18) V → aux verb
- 19) V → aux not verb

非終端記号 { SS: 文章
S: 文
NP: 名詞句
VP: 動詞句
NP1: 名詞句1
PP: 前置詞句
ADJP: 形容詞句
V: 述語

終端記号 { end: ピリオド
conj: 接続詞
det: 限定詞
noun: 名詞
pron: 代名詞
verb: 動詞
aux: 助動詞
not: 否定語
adj: 形容詞
prep: 前置詞

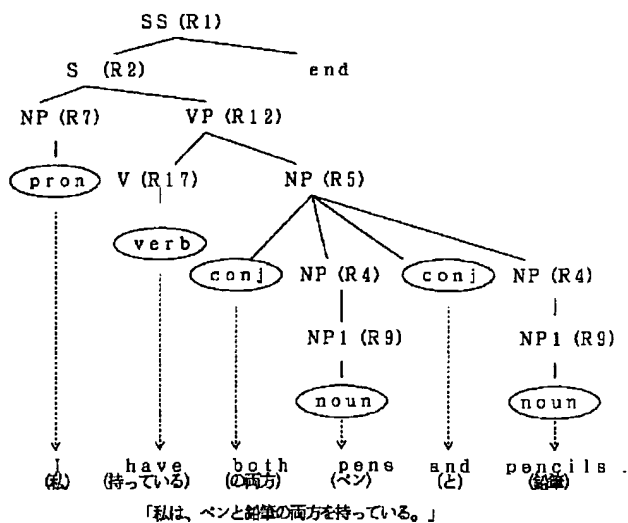
【図3】



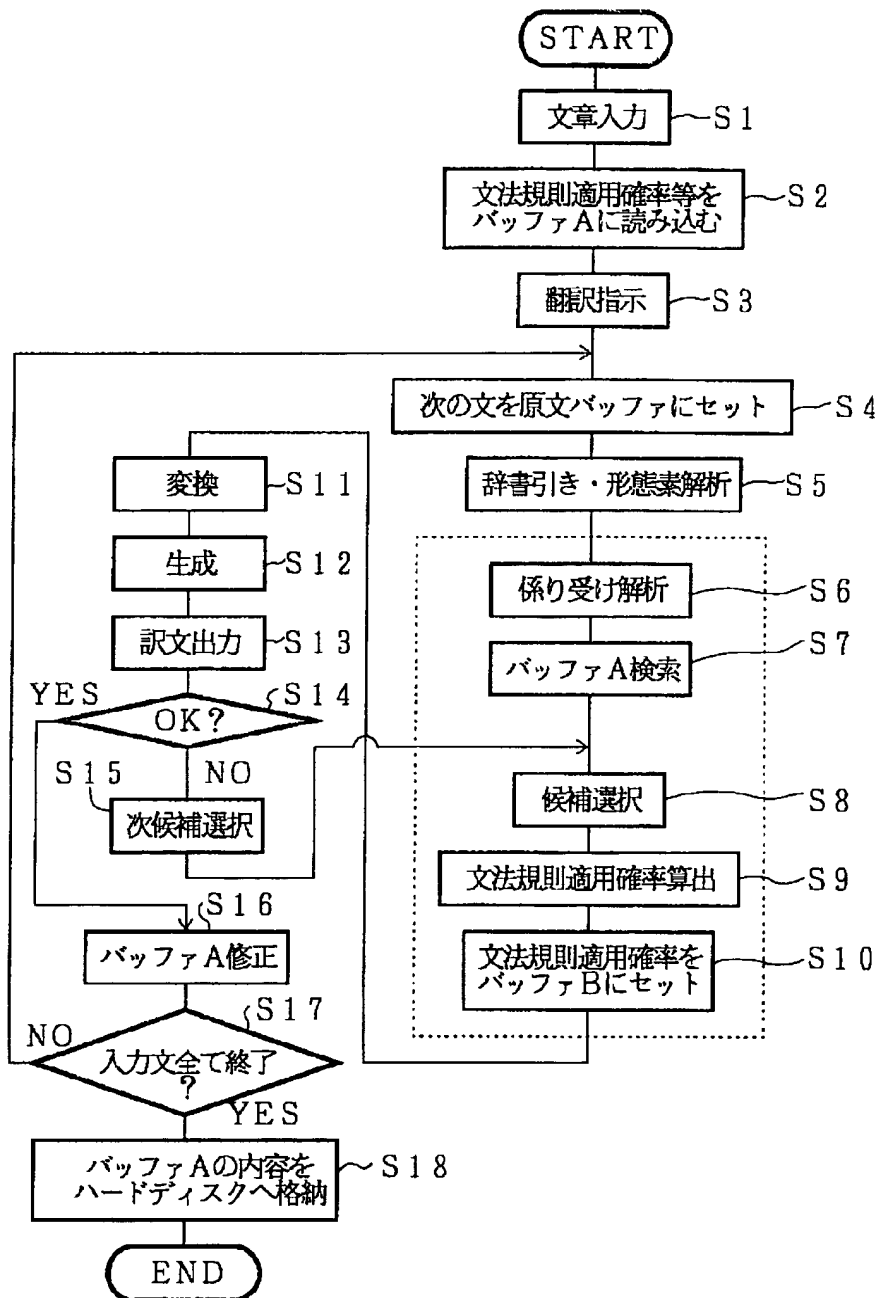
【図6】

文法規則番号	適用回数	該当非終端記号 の出現回数	文法規則適用確率
1	1	1	1
2	1	1	1
3	1	6	0.16666666666666666
4	1		0.16666666666666666
5	1		0.16666666666666666
6	1		0.16666666666666666
7	1	2	0.5
8	1		0.5
9	1	4	0.25
10	1		0.25
11	1		0.25
12	1		0.25
13	1	3	0.3333333333333333
14	1		0.3333333333333333
15	1	1	1
16	1	1	1
17	1	3	0.3333333333333333
18	1		0.3333333333333333
19	1	1	1

【図4】



【図5】



【図12】

	1	have	both	pens	and	pencils	,
1	noun	verb	det	noun	conj	noun	end
2			conj				
3			pron				
↑							

品詞優先度

【図7】

文法規則番号	適用回数	該当非終端記号 の出現回数	文法規則適用確率
1	1	1	1
2	1	1	1
3	1	}	0.25
4	1		0.25
5	0		0.25
6	1		0.25
7	1	}	0.25
8	0		0.25
9	2	}	1
10	0		0
11	0	}	0
12	1		0
13	0		0
14	0		0
15	0	0	0
16	0	0	0
17	1	}	1
18	0		0
19	0		0

【図8】

文法規則番号	適用回数	該当非終端記号 の出現回数	文法規則適用確率
1	1	1	1
2	1	1	1
3	0	}	0.50
4	2		0.25
5	1		0.25
6	0		0.25
7	1	}	0.25
8	0		0
9	2	}	1
10	0		0
11	0	}	0
12	1		0
13	0		0
14	0		0
15	0	0	0
16	0	0	0
17	1	}	1
18	0		0
19	0		0

【図9】

文法規則番号	適用回数	該当非終端記号 の出現回数	文法規則適用確率
1	2	2	1
2	2	2	1
3	1	}	0.10
4	2		0.20
5	2		0.20
6	1		0.10
7	2	}	0.20
8	1		0.10
9	3	}	0.75
10	1		0.25
11	1	}	0.20
12	2		0.40
13	1		0.20
14	1		0.20
15	1	1	1
16	1	1	1
17	2	}	0.50
18	1		0.25
19	1		0.25

【図11】

文種	文法規則番号	適用回数	該当非終端記号 の出現回数	文法規則適用確率
なし	1	1	1	1
	2	1	1	0.25
	3	1	1	0.25
	4	1	1	0.25
	5	1	1	0.25
マニュアル	1	1	1	1
	2	1	1	0.50
	3	1	1	0.50
	4	1	1	0.50
	5	1	1	0.50
新聞	1	1	1	1
	2	1	1	0.16
	3	1	1	0.16
	4	1	1	0.16
	5	1	1	0.16
...

【図10】

